Forecasting Electricity Spot Prices Accounting for Wind Power Predictions

Tryggvi Jónsson, Pierre Pinson, *Member, IEEE*, Henrik Aa. Nielsen, Henrik Madsen, Torben S. Nielsen

***Abstract*—A two-step methodology for forecasting of electricity spot prices is introduced, with focus on the impact of predicted system load and wind power generation. The non-linear and non-stationary influence of these explanatory variables is accommodated in a first step based on a non-parametric and time-varying regression model. In a second step, time-series models i.e. ARMA and Holt-Winters, are applied to account for residual autocorrelation and seasonal dynamics. Empirical results are presented for out-of-sample forecasts of day-ahead prices in the Western Danish price area of Nord Pool's Elspot, during a two year period covering 2010-2011. These results clearly demonstrate the practical benefits of accounting for the complex influence of these explanatory variables.**

***Index Terms*—Electricity prices, nonlinear modeling, nonparametric modeling, forecasting, adaptivity, robustness**

I. Introduction

THE PARTICIPANTS in deregulated electricity markets rely, among other things, on forecasts of future prices for bidding and optimizing the dispatch of their generation units. Methods for deriving such forecasts can be divided into two fundamentally different categories: economical equilibrium models that mimic the actual pricing model, and statistical ones. The former models are able to provide excellent forecasts when given sufficiently accurate information (see e.g. [1], [2] and references therein). This information is however seldom available to individual market participants. In addition, the presence of non-dispatchable yet cheap generation units in the system implies that this information might be impossible to obtain since their production is indeed stochastic. Statistical approaches then appear to be a relevant alternative.

The increased focus on curbing carbon emissions worldwide has led to vast investments in renewable energy sources and in particular wind power. Many of these emerging energy sources, wind power included, share a characteristic in being non-dispatchable due to the varying availability of the fuel, which also cannot be stored. Consequently these sources are ill-suited for long term contracts, leaving only markets with relatively short time between gate-closure and delivery as a realistic option for selling the production. Inevitably the prices at these markets are affected by this additional supply [3]–[7].

The impact of renewable energy is superimposed on already existing price features such as non-stationarity, periodicity, mean reverting spikes, positive skewness and high kurtosis along with intra- and inter-day serial correlation [8]–[10]. These features arise from the distinct characteristics of electricity as a commodity. Firstly, lack of direct storability along with the specialized and technically limited transmission system required, makes arbitrage over time and space difficult [11], [12]. Secondly, demand for electricity is highly inelastic while exhibiting strong seasonalities in the short term. Meanwhile the supply function is discontinuous, convex and steeply increasing at the high production end [9], [13], [14]. The aim of the present paper is to propose a forecasting methodology which allows for accommodating the effect of the emerging renewable sources as well as the characteristics described in [8]–[10], [14].

In [3], predicted power production is shown to significantly impact the distribution moments of day-ahead electricity prices through a Danish case study. Motivated by these findings, the present paper introduces a two-step methodology for issuing point forecasts for electricity spot prices, accounting for the impact of predicted load and wind power production. First, a time-varying function is estimated, jointly mapping the predicted hourly load and wind power production to a corresponding spot price. The function is built based on a conditional parametric regression model for which the parameters are estimated adaptively. In addition, past observations are discounted exponentially as new ones become available. The resulting flexibility in the model serves the purpose of accommodating both the non-linear relationship between the explanatory variables and the prices, as well as the non-stationarity of all the processes involved. Although here a conditional parametric model is chosen for the inclusion of the wind power and load as explanatory variables, other model types might be just as suitable. For instance including the forecast wind power production along with the load in the adaptive wavelet neural-network model of [15]. Regardless of the model chosen, the main message of this paper remains intact: that wind power production, where present, impacts the prices to such extent that it should be accounted for in a forecasting model for electricity spot prices.

In the second stage, the remaining residuals are modeled using well known models from the time series analysis literature. These models are an additive double seasonal Holt-Winters model and a recursively estimated seasonal AR model. All models are estimated under robust criteria in order to protect the parameter estimates from the effects of excessive price spikes. Models and parameters are optimized in terms of weighted least squares residuals.

The sole focus of this paper is deriving a model that

T. Jónsson is with ENFOR A/S, Hørsholm, Denmark and the Department of Informatics and Mathematical Modelling at the Technical University of Denmark, Kgs. Lyngby, Denmark (e-mail: tj@enfor.dk)

P. Pinson and H. Madsen are with the Department of Informatics and Mathematical Modelling at the Technical University of Denmark.

H. Aa. Nielsen and T.S. Nielsen are with ENFOR A/S

describes the expected spot prices on a day-ahead basis. Although the authors recognize proper modeling of forecast uncertainty and price spikes as paramount in forecasting electricity spot prices, the empirical features of the prices are such that appropriate uncertainty or spike modeling will easily comprise a full paper of its own. The model presented in this paper can however be combined with a price spike model (e.g. those presented in [16], [17]). Same goes for a model for predictive densities such as the ones described in [9] and [18].

The context of the empirical results presented in this paper is the Western Danish price area (DK-1) of the Nord Pool's Elspot market. Operational data for the period from November 1st 2008 until December 31st 2011 are considered and used to evaluate the model's day-ahead forecasting skill. Furthermore, the value of the time-adaptivity and robustness is illustrated by comparing the performance of the proposed approach against its time-invariant and non-robust counterparts. Despite focus being on this market only, the fact that results on the influence of wind power forecasts on electricity prices similar to those of [3] have been obtained for other areas, e.g. Germany and Spain [5], indicates that similar forecasting methodology could be applied successfully in the context of other markets.

The remainder of the paper is structured as follows. The market and data on which the empirical work is based are described in Section II. Section III presents the models and Section IV the obtained empirical results. Finally concluding remarks are given in Section V.

II. Empirical Background

Elspot is a day-ahead market for physical delivery of electric power, operated by Nord Pool Spot AS [19] in the entire Scandinavia (Denmark, Finland, Norway and Sweden) and in Estonia.

On Elspot, contracts for next day physical delivery are traded for hourly periods. Prices are set as the intersection between the aggregated supply and demand curves for each hour of the day, right after gate-closure at noon. The intersection of the curves representing all bids in the entire market region defines the system price. The system price, in addition to serve as reference for financial contracts, is the price at which physical contracts are settled if transmission capacity is sufficient throughout the entire region.

Due to limited transmission capacity however, both between and within the member countries, the market region is divided into several price areas. If the scheduled flow between price areas exceeds the corresponding transmission capacity, area prices that differ from the system price are calculated. On such occasions, the area prices are identical among areas that have sufficient capacity on their interconnections. Areas on each side of a congested connection however have different prices. The area prices are the ones at which contracts for physical delivery are settled.

The context of the empirical results presented in the following is the Western Danish price area (DK-1) of Elspot. The area comprises Jutland and Funen along with the islands west of the Størebælt channel and has relatively strong connections to both Norway and Sweden to the north and Germany to the south. As of August 20, 2010 the DK-1 area also has a 600 MW link to the Eastern Danish price area (DK-2). Furthermore, the area has a large share of its annual electricity consumption (about 25%) generated by wind turbines.

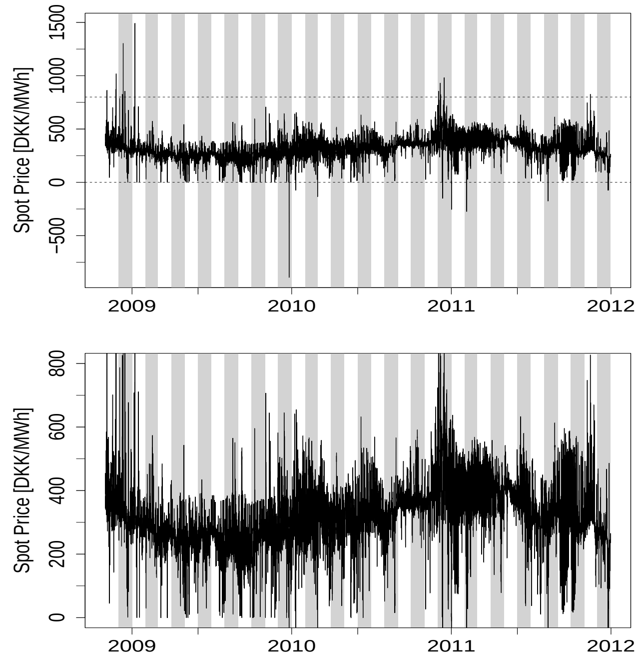

Fig. 1. Time series plots of the spot prices throughout the considered period.

The data set used consists of hourly observed area prices along with forecasts of both wind power production and consumption in the area. Both forecasts are issued before gate-closure and have a temporal resolution of 1 hour. The observed prices are taken from the website of the Danish transmission system operator (TSO), Energinet.dk ([20]). The load forecasts are the ones made publicly available by Nord Pool through their website ([19]). The wind power production forecasts however stem from a statistically based wind power prediction software [21]. Time series plots of the prices for the considered period are shown in Fig. 1. Whereas the top plot shows the full series, the bottom one shows the sereis on a scale truncated a 0 and 800 DKK/MWh.

III. Model for Spot Prices

A. The Rationale Behind the Proposed Modeling Approach

As more than half of the annual electricity production in the Nordic region is hydro power based [22], the prices at the Elspot market are inevitably dominated by the water stock in the hydro power reservoirs in Norway, Sweden and Finland. This stock however varies relatively slowly compared to the resolution and lead-times of the desired forecasts. Indeed the fact that data for these are published with a resolution of one week should be enough to convince one that such data has no explanatory value in day-ahead price forecasts with a resolution of a single hour. Instead, the impact of the water stock appears as a slow drift in the price series with this resolution. Because of this, it is decided only to implicitly include the impact of hydro stock, along with other slowly varying fundamentals such as fuel prices, by adaptive estimation of the model parameters.

In [3] the ratio between predicted wind power production and forecast consumption is shown to affect the area spot prices in DK-1 substantially. On a similar note, forecast wind power production is shown to appear in the supply function as a stochastic threshold in [4]. The reason for wind power to have such a strong influence on prices is owed to how it fundamentally differs from most other energy sources that significantly contribute to the supply. Whereas conventional power plants can be scheduled to steadily produce a certain amount of energy over a longer period, wind turbines literally produce as the wind blows. In addition, the absence of fuel costs allows wind turbines to produce at a marginal cost close to 0. This low marginal cost causes wind power to enter the supply function to the very left and thereby horizontally shifts the supply function.

The stochastic fuel availability moreover implies that those responsible for bidding wind power into the market have to rely on production forecasts for their decisions. As a consequence, the supply function comprises the production forecasts available at gate-closure and not the realized volume. For this reason, forecasts of future production are the appropriate form of wind power for any inference on its relation to the prices. This aside, forecasting in practice has to be based on predictions of both wind power production and load. Thus following [23] the model is entirely and based on forecast values for the explanatory variables.

In contrast to [3] the two explanatory variables enter the former model step individually and not as a ratio. This is because that formulation was found to yield better forecasts of the prices.

In [35] lagged values of measured load are found to have significant explanatory power in a model for the spot prices. Past demand has however no direct effect on the current spot prices. It is therefore likely that the effect seen in [35] is owed to that in the absence of actual load forecasts the lagged demands serve as implicit prediction of the load. The high number of lags, especially seasonal ones, found significant in [35] support this conclusion. Thus with access to an actual load forecast, no attempt to include past demand in the model was made.

B. Spot Price as a Function of Forecast Wind Power Production and Load

An excellent general description of the methodology and estimation procedure used to describe the spot prices as a function of wind power and load forecasts is given in [24] (without recursivity) and in [25], [26](including recursivity and robustness respectively). However, in order to make this paper self-contained, an outline of the method is given here, tailored to the application at hand.

Let a model for the spot price at time t, $\pi_t^{(S)}$, be denoted as

$$\pi_t^{(S)} = \boldsymbol{\theta}_t(\boldsymbol{u}_t) + \varepsilon_t \tag{1}$$

where $\boldsymbol{\theta}_t(\cdot)$ is a function of a set of explanatory variables, $\boldsymbol{u}_t$, and ε_t is a noise term, centered and a with a finite variance. Thus, the model (1) is a non-linear and non-parametric regression model.

The function $\boldsymbol{\theta}(\cdot)$, is approximated using polynomials by fitting a linear model at a number of distinct *fitting points*. More specifically let $\boldsymbol{u} = [\ u_1 \quad u_2\ ]^T$ denote a particular fitting point and let $\boldsymbol{p}_d(\boldsymbol{u})$ denote a column vector containing the terms in the corresponding polynomial of order d. Here $d = 2$ has been chosen after trials with $d \in \{1, 2, 3\}$ yielding $\boldsymbol{p}_2(\boldsymbol{u}) = [\ 1 \quad u_1 \quad u_2 \quad u_1^2 \quad u_1u_2 \quad u_2^2\ ]^T$.

Now define

$$\boldsymbol{\phi}_{u,t}^T = [\ \phi_{u_1,t} \quad \ldots \quad \phi_{u_2^2,t}\ ] \tag{2}$$

a column vector of coefficients such that the model

$$\pi_t^{(S)} = \boldsymbol{p}_2^T(\boldsymbol{u}_t)\boldsymbol{\phi}_{u,t} + e_t \tag{3}$$

describes the prices in the close vicinity of the fitting point $\boldsymbol{u}$ where e_t is a noise term, centered and with a finite variance.

The parameters in (3) are estimated using recursive and robust weighted least squares. That is

$$\widehat{\boldsymbol{\phi}}_{u,t} = \arg\min_{\boldsymbol{\phi}_{u,t}} \sum_{s=1}^{t} \lambda^{t-s} w_u(\boldsymbol{u}_s) \left(g\left(e_s, \tau\right)\right)^2 \tag{4}$$

where $e_t = \pi_t^{(S)} - \boldsymbol{p}_2^T(\boldsymbol{u}_t)\boldsymbol{\phi}_{u,t}$ and $0 < \lambda < 1$ is a forgetting factor that exponentially discounts observations over time. Furthermore, $w_u(\boldsymbol{u}_t)$ is a weight, assigned to observation $\boldsymbol{u}_t$ as a function of its distance to the fitting point $\boldsymbol{u}$. Finally, $g(\cdot, \cdot)$ is the *Huber influence function* [27], defined as

$$g(e_t, \tau) = \mathrm{sgn}(e_t) \cdot \min\left\{|e_t|, \tau\right\}, \tag{5}$$

where τ is the cut-off value or the maximum influence a single observation is allowed to have on the estimate.

The weights are assigned as

$$w_u(\boldsymbol{u}_t) = W\left(\frac{||\boldsymbol{u}_t - \boldsymbol{u}||}{h(\boldsymbol{u})}\right) \tag{6}$$

where $W(\cdot)$ is a function taking non-negative arguments, $||\cdot||$ denotes the Euclidean norm and $h(\boldsymbol{u})$ is the bandwidth applied in the fitting point $\boldsymbol{u}$. Following [24] and [25] a tri-cube kernel is used to determine the weights. That is

$$W(x) = \begin{cases} (1 - x^3)^3 & \text{if } x \in [0; 1) \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

which entails weights between 0 and 1.

It can be shown (see e.g. [26] or [28], Ch. 11) that the adaptive parameter estimates in Eq. (4) can be found recursively as

$$\widehat{\boldsymbol{\phi}}_{u,t} = \widehat{\boldsymbol{\phi}}_{u,t-1} + w_u(\boldsymbol{u}_t)\boldsymbol{R}_{u,t}^{-1}\boldsymbol{p}_2(\boldsymbol{u}_t)g\left(e_{t|t-1}, \tau\right) \tag{8}$$

where

$$e_{t|t-1} = \pi_t^{(S)} - \boldsymbol{p}_2^T(\boldsymbol{u}_t)\widehat{\boldsymbol{\phi}}_{u,t-1} \tag{9}$$

and

$$\boldsymbol{R}_{u,t} = \lambda\boldsymbol{R}_{u,t-1} + w_u(\boldsymbol{u}_t)\frac{\partial g(e_t, \tau)}{\partial e_t}\boldsymbol{p}_2(\boldsymbol{u}_t)\boldsymbol{p}_2^T(\boldsymbol{u}_t). \tag{10}$$

Abruptly changing parameter estimates are avoided by following [25] and defining the effective forgetting factor, λ_t^* as

$$\lambda_t^* = 1 - (1-\lambda) w_u(\boldsymbol{u}_t) \frac{\partial g(e_t, \tau)}{\partial e_t} \tag{11}$$

and subsequently update (10) so it becomes

$$\boldsymbol{R}_{u,t} = \lambda_t^* \boldsymbol{R}_{u,t-1} + w_u(\boldsymbol{u}_t) \frac{\partial g(e_t, \tau)}{\partial e_t} \boldsymbol{p}_2(\boldsymbol{u}_t) \boldsymbol{p}_2^T(\boldsymbol{u}_t). \tag{12}$$

Finally, $\boldsymbol{\theta}_t(\boldsymbol{u})$ is estimated by

$$\widehat{\boldsymbol{\theta}}_t(\boldsymbol{u}) = \boldsymbol{p}_2^T(\boldsymbol{u}) \widehat{\boldsymbol{\phi}}_{u,t} \tag{13}$$

and estimates for other values of $\boldsymbol{u}_t$ than the fitting points are found by linear interpolation.

In contrast to fitting 24 hour-specific models, a single conditional parametric model is estimated for all hours of the day simultaneously. The rationale behind this choice is twofold. First the apparent diurnal seasonality in the prices is mainly caused by that of the demand. Thus the seasonality is implicitly accounted for by the inclusion of the load forecast as an explanatory variable. Secondly, the consumption pattern is in most cases quite similar among consecutive hours. Thus, fitting hour-specific model in many cases leads to the exclusion of observations of similar circumstances from neighboring hours. The absence of obvious regime shifts in the consumption pattern makes alternative segmentation also problematic. Besides, all data split results in longer time passing between observations prompting a lower forgetting factor and thereby less stable parameters over time. So the dynamics of the spot price most local in time along with seasonalities not owed to the demand and wind are left to be accommodated in the second model step. A consequence of adopting this fitting procedure is that the results from the former step are only to be viewed for model building purposes and not evaluated on their own. This is because the missing diurnal variation in the function will inflate the performance measures.

For estimation, the independent variables, i.e. forecast wind power and load, are scaled such that $u_i \in [-1,1] \forall i$ using the range of each variable in the training set to perform the scaling. Fitting points are then chosen as 24 equidistant ones in each dimension. It was decided not to optimize neither the position of the fitting points nor their number since results from a few different sets of fitting points indicated that little would be gained from their inclusion in the optimization. Such optimization is however possible, e.g. by methods presented in [29] and [36].

The model parameters are estimated using a nearest neighbor bandwidth which implies that the actual bandwidth varies with the local density of the data. That is, the bandwidth for each fitting point is chosen such that a certain fraction γ of the observations fulfill $||\boldsymbol{u}_s - \boldsymbol{u}|| \leq h(\boldsymbol{u})$. The actual bandwidth for each fitting point is found empirically from the training set. Put differently, the bandwidth for each particular fitting point is set as the γ-quantile of the Euclidean distances between that fitting point and the observations in the training set.

The actual values of γ, λ and τ are selected by a least squares optimization of the forecasts issued at noon the day before delivery. More precisely let $\widehat{\pi}_{t+k|t}^{(S)}$ denote the forecast spot price for time $t+k$ issued at time t. Every day at noon, forecasts are issued for the period from midnight to midnight the following day and then no forecasts are made until noon the next day, when forecasts for the same lead times are generated. This implies that forecasts for individual hours of the day always have the same lead time. This scenario resembles the practical one and these forecasts are termed *day-ahead forecasts* in the following and noted as $\widehat{\pi}_{DA(t)}^{(S)}$. This notation implies that for an observation $\pi_t^{(S)}$ the corresponding day-ahead forecast is $\widehat{\pi}_{DA(t)}^{(S)} = \widehat{\pi}_{t|t-13}^{(S)}$ if t corresponds to the first hour of the day. If the observation is from the second hour of the day, $\widehat{\pi}_{DA(t)}^{(S)} = \widehat{\pi}_{t|t-14}^{(S)}$, and so forth. The optimal values of the tuning parameters are then found as

$$\begin{bmatrix} \gamma^* \\ \lambda^* \\ \tau^* \end{bmatrix} = \underset{\gamma,\lambda,\tau}{\arg\min}\, \mathrm{RMSE}_{DA}(\gamma, \lambda, \tau). \tag{14}$$

where

$$\mathrm{RMSE}_{DA}(\gamma, \lambda, \tau) = \sqrt{\frac{1}{N} \sum_{t=1}^{N} \left(\pi_t^{(S)} - \widehat{\pi}_{DA(t)}^{(S)} \right)^2}. \tag{15}$$

C. Residual models

The purpose of the model's second step is to account for autocorrelation and seasonal patterns that are not explained by the load and the wind power. Out of the models tried for the second step of the model, two models turned out to be superior to the others and yet quite compatible. These models are seasonal AR model with robust and adaptively estimated parameters, and a seasonal additive Holt-Winter model, also estimated under a robust criteria.

The autocorrelation function (ACF) and the partial autocorrelation function (PACF) for the residuals from the first model step are shown in Figure 2. Despite that the residual series is not completely stationary, its ACF and PACF are used to identify potentially appropriate orders of AR and MA terms to include in second step model. Afterwards a survey of different model orders is conducted in order to determine the most appropriate structure. From this survey, a model on the form

$$\varepsilon_{t+k} = \boldsymbol{z}_t^T(k) \boldsymbol{\beta}_t(k) + v_{t+k} \tag{16}$$

is found to be appropriate for $k \leq 24$. The vectors $\boldsymbol{z}_t(k)$ and $\boldsymbol{\beta}_t(k)$ are defined as

$$\boldsymbol{z}_t(k) = \begin{matrix} [& 1 & \varepsilon_{t-1} & \varepsilon_{t-2} & \cdots \\ & \varepsilon_{t+k-24} & \varepsilon_{t+k-48} & \varepsilon_{t+k-168}]^T \end{matrix} \tag{17}$$

$$\boldsymbol{\beta}_t(k) = \begin{bmatrix} \beta_{0,t,k} & \beta_{1,t,k} & \ldots & \beta_{6,t,k} \end{bmatrix}^T, \tag{18}$$

where in turn $\beta_{j,t,k}$ are parameters to be estimated recursively. Moreover, ε_t is defined by Eq. (1) and v_t is a new noise term also centered and with finite variance. Put differently, separate model parameters are estimated for each lead time, relevant for a day-ahead forecast, that correspond to the lagged values of the forecast error from the first model step (ε). Obviously for $k = 23$, $\varepsilon_{t-1} = \varepsilon_{t+k-24}$ and correspondingly for $k = 24$, $\varepsilon_t = \varepsilon_{t+k-24}$. In these special cases, the dimension of the

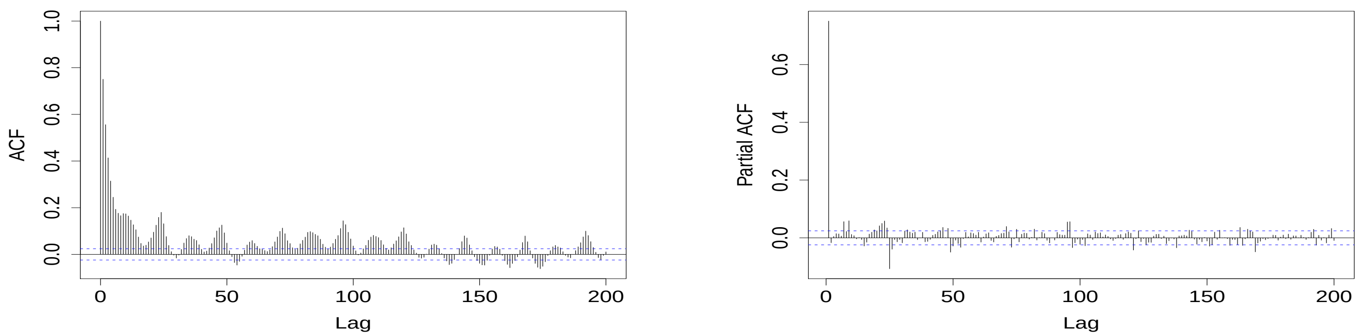

Fig. 2. ACF (left) and PACF (right) for the residuals arising from the first model step.

design matrix is reduced so that each observation is only represented once. Other model structures, such as including a moving average term were considered but the one here describes was found to be the most appropriate one.

The parameter estimates are obtained similarly to what already has been described for the first step of the model or as

$$\widehat{\boldsymbol{\beta}}_t = \arg\min_{\boldsymbol{\beta}_t} \sum_{s=1}^{t} \lambda^{t-s} \left(g(v_t, \tau)\right)^2 . \tag{19}$$

where $v_t = \varepsilon_t - \boldsymbol{z}_t^T \boldsymbol{\beta}_{t-1}$, $g(\cdot, \cdot)$ is defined by Eq. (5) and $0 < \lambda < 1$ is a forgetting factor as before. Hereafter, the AR model with these parameter estimates will be referred to as RLS-AR. The difference between the two procedures is merely that the kernel weights, $w_u(\cdot)$, are omitted from all the equations so that Eq. (8) and (12) become

$$\widehat{\boldsymbol{\beta}}_t = \widehat{\boldsymbol{\beta}}_{t-1} + \boldsymbol{R}_t^{-1} \boldsymbol{z}_t g\left(v_t, \tau\right) \tag{20}$$

$$\boldsymbol{R}_t = \lambda_t^* \boldsymbol{R}_{t-1} + \frac{\partial g(v_t, \tau)}{\partial v_t} \boldsymbol{z}_t \boldsymbol{z}_t^T . \tag{21}$$

respectively, where

$$\lambda_t^* = 1 - (1 - \lambda) \frac{\partial g(v_t, \tau)}{\partial v_t} . \tag{22}$$

The Holt-Winters model was initially introduced in [30] for one seasonal cycle while extension to multiple cycles is described in [31]. The model that eventually yielded the best prediction skill for ε_t only has a single daily seasonal cycle. However since the benchmark model for $\pi_t^{(S)}$ was found to benefit substantially from including a weekly seasonality as well, a formulation for a double seasonal model is given here. The transition from a double seasonal model to a single seasonal one merely involves omitting the second seasonality from all equations.

The purely additive form of the Holt-Winters model is used (see e.g. [32] for a comparison between additive and multiplicative Holt-Winters models). The model contains a mean term, μ_t, and two separate seasonal indices, D_t and W_t. The period of D_t is 24 while that of W_t is 168, corresponding the within-day and within-week seasonalities respectively. A standard non-robust Holt-Winters model can be denoted as

$$\mu_t = \alpha_\mu \left(\varepsilon_t - (D_{t-24} + W_{t-168})\right) + \left(1 - \alpha_\mu\right) \mu_{t-1} \tag{23}$$

$$D_t = \alpha_D \left(\varepsilon_t - (\mu_t + W_{t-168})\right) + \left(1 - \alpha_D\right) D_{t-24} \tag{24}$$

$$W_t = \alpha_W \left(\varepsilon_t - (\mu_t + D_{t-24})\right) + \left(1 - \alpha_W\right) W_{t-168} \tag{25}$$

where the α's are smoothing parameters to be estimated. Once the different terms of the model are updated, the k-step ahead forecast is found as

$$\widehat{\varepsilon}_{t+k|t} = \mu_t + D_{t+k-24} + W_{t+k-168} . \tag{26}$$

The inclusion of a trend term in the model was considered but the resulting improvement in forecasting skill was found to be insignificant.

Writing Eq. (23) - (25) on their error correction form and adopting the formulae for robustness from [33] yields

$$\mu_t = \mu_{t-1} + \alpha_\mu g(v_t, \tau) \tag{27}$$

$$D_t = D_{t-24} + \alpha_D g(v_t, \tau) \tag{28}$$

$$W_t = W_{t-168} + \alpha_W g(v_t, \tau) \tag{29}$$

where, as before, $v_t = \varepsilon_t - \widehat{\varepsilon}_{t|t-1}$ and $g(\cdot, \cdot)$ is the Huber influence function given by Eq. (5).

For both models, a single set of tuning parameters was estimated for all lead times. As for the first step, the parameters are optimized with respect to the day-ahead RMSE as formulated in Eq. (14) and (15). Certainly these parameters could be optimized for each hour of the day. A search for initial values indicated however, that improvement in forecasting skill achieved by doing so would only be marginal. This choice does not in any way alter the validity of the model and the results obtained but only indicates that some of the parameters might by slightly sub-optimal.

IV. Empirical Results

The parameters in the two model steps are estimated sequentially based on the first 14 months of the data set or from November 2008 and through December 2009. The remaining two years of data are then used as an independent test period. For estimation, prices above 800 DKK/MWh and below 0 DKK/MWh are excluded to avoid unstable parameters. Performance estimates presented in the following are based on all observations though.

In order to illustrate the contribution from different features of the model, different reference models are estimated and their RMSE and Mean Absolute Error (MAE) compared to that of the proposed model. The RMSE and the MAE are both presented in two versions:

1) On the price's real scale (in DKK), and

TABLE I
ESTIMATED PARAMETERS AND FORECASTING SKILL FOR THE FIRST MODEL STEP AND THE REFERENCE MODELS

Explanatory Variables(s)	Estimation Setup	γ	λ	τ	In-sample RMS(S)E [DKK]	In-sample RMS(S)E [−]	In-sample MA(S)E [DKK]	In-sample MA(S)E [−]	Out-of-sample RMS(S)E [DKK]	Out-of-sample RMS(S)E [−]	Out-of-sample MA(S)E [DKK]	Out-of-sample MA(S)E [−]
Wind Power & Load Forecasts	Recursive & Robust	0.8529	0.9877	55.67	51.10	0.663	28.33	0.705	50.89	0.676	35.05	0.732
	Recursive only	0.9018	0.9901	—	52.33	0.679	29.84	0.742	51.85	0.686	35.52	0.742
	Time-invariant	0.0821	—	—	52.05	0.676	32.07	0.798	101.54	1.349	86.17	1.799
Load Forecasts	Recursive & Robust	0.7887	0.9831	53.38	56.24	0.730	30.66	0.763	58.79	0.781	39.09	0.816

2) as a skill relative to the daily persistence (RMSSE and MASE).

That is, the measures are scaled by the corresponding measures for a daily persistence forecast as suggested by [34]. More formally, the RMSE is scaled by

$$\sqrt{\frac{1}{N_{per}-24}\sum_{t=25}^{N_{per}}\left(\pi_t^{(S)}-\pi_{t-24}^{(S)}\right)^2} \tag{30}$$

and the MAE is scaled by

$$\frac{1}{N_{per}-24}\sum_{t=25}^{N_{per}}\left|\pi_t^{(S)}-\pi_{t-24}^{(S)}\right| \tag{31}$$

where N_{per} is the number of observations in the sample for which the measure is calculated. This scaling yields a relative error measure that is unbiased towards forecasting ability of high and low prices and does not call for any data trimming due to the prices being zero or close to that. A more detailed discussion on the RMSSE, MASE and forecast accuracy measures in general can be found in [34].

For the first model step the in-sample and out-of-sample performance is compared to that of:

1) its time-invariant and non-robust counterpart,
2) its non-robust counterpart,
3) a model estimated in the same manner but only taking load as an explanatory variable.

Finally, the forecasting skill of the combined models is compared to the that of

1) two seasonal persistence models, one with a daily period, and another with a weekly period,
2) the previously described Holt-Winters and RLS-AR models applied directly to the spot price series.
3) a series of 24 ARIMAX models, one for each hour of the day with the forecast wind and load as external regressors.

In line with such type of models in the existing literature (e.g. [8], [35]) the ARIMA models are fit in terms of $log(\pi_t^{(S)}+1000)$. The model order for each hour is decided on by minimizing the Bayesian Information Criteria (BIC) [36] for the training period. The external variables are considered both on their original scale and log-transformed as suggested by [35]. After calculating the predictions, they are transformed to the original scale by the exponential of the prediction plus half the estimated variance. The set of external variables that yielded the best forecasting skill was the one with the log-transformed wind power and load forecasts for which results are reported in the following.

The fitting points for the former model step are chosen as 24 equidistant ones in each dimensions thus yielding a grid of 24^2 equidistant fitting points in total. For each point, the coefficient vector, $\boldsymbol{\phi}$, is initialized by setting all its elements to 0.1. The corresponding matrix inverse variance-covariance matrix, $\boldsymbol{R}_0$, is chosen as a diagonal one with non-zero elements as 10^{-6}. Thereafter the first 1008 observations are taken for initialization and are excluded from the performance measure. Hence the tuning parameters are optimized in context of the previously mentioned training period apart from its first 42 days. During the initialization period, the robust criteria is relaxed in order to obtain frequent updates of $\boldsymbol{R}_t$. This is necessary because of the poor initial guesses for the coefficients.

The parameters are optimized as described in the previous section, using the quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) method. The estimated parameters are shown in Table I along with the corresponding in- and out-of-sample RMSE.

The results indicate that the inclusion of wind power forecasts and the recursive parameter estimation are worth the effort since the top two models significantly outperform the bottom two. Especially the recursive parameter estimation seems to paramount since the performance of the time invariant model degrades excessively during the test period. In terms of performance, the benefits of the robust estimation are less obvious. However, given the spiky behavior of the spot prices, it is generally sound to robustify the estimation process in order to protect the model from abrupt changes caused by a single spike. In light of the varying volatility of the prices, making τ recursive, as described in [26] and [37], could be more appropriate. No such efforts were made for this paper though.

A forgetting factor of $\lambda = 0.9877$ translates to that $1/(1-0.9877) = 81.3$ latest observations are effective in the parameter estimation which corresponds to around 3.5 days. Owed to the locally weighted and robust estimation however, the effective forgetting factor, $\lambda^{(*)}$ is somewhat higher and varies between fitting points. In [25] a procedure to estimate the actual memory of the model is proposed. Following this procedure and averaging over time as well

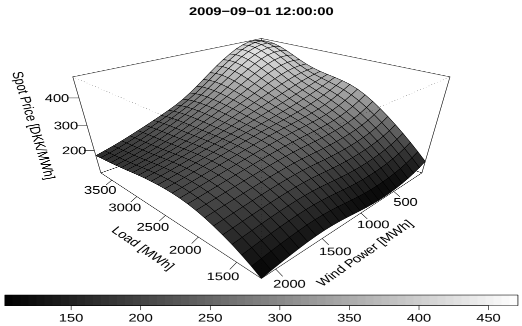

Fig. 3. Surface plot of the spot prices as a function of the forecast wind power production and load at noon on September 1st 2009

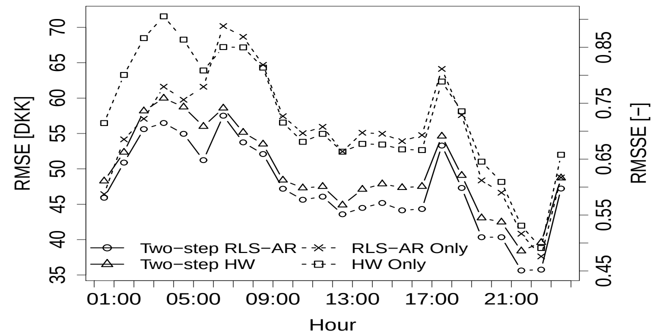

Fig. 4. RMSE for individual lead times during the test period in DKK (left axis) and as percentages of the average price for the period (right axis).

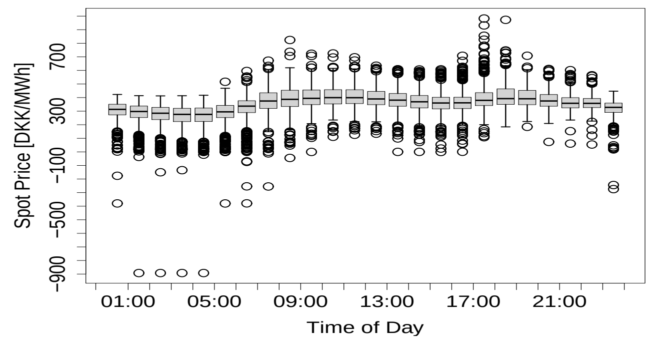

Fig. 5. An hour-by-hour box plot of the spot prices during the test period. Whiskers are placed at 1.5 times the interquartile range.

as the 576 fitting points yields a mean number of effective observations, $\overline{\eta} = 281.8$ hours and a corresponding average effective forgetting factor of $\overline{\lambda}^{(*)} = 0.9965$.

The function which the model approximates is shown for one instance in Figure 3. This function is then updated each day at noon by adjusting the model's coefficients to observations from the current day. Subsequently, forecasts are calculated from next day's input forecasts, i.e. the wind power and the load forecast. This is done by bilinear interpolation between the four fitting points surrounding the input forecasts. In other words, the price forecast for a set of input forecasts for a given hour, $\boldsymbol{u}_{t+k}$, is found as the corresponding point on the linear plane joining the four nearest fitting points. Alternatively generalization to values other than fitting points could be done non-linearly, e.g. by one of the multidimensional spline techniques presented in [36]. In light of the high number of fitting points used here, linear interpolation was however deemed adequate.

The initial values for the coefficients, $\boldsymbol{\beta}_0$, in the second step AR model are found from a standard AR model, i.e. not recursively estimated, using observations for the first 42 days of the data set. The inverse of the corresponding variance-covariance matrix is taken as $\boldsymbol{R}_0$. For the Holt-Winters model, an initial value for the μ-term (μ_0) is found as the mean price during the first 42 days of the training period. The seasonal terms are initialized as the difference between μ_0 and the average price for the individual hours during the same 42 days. In the same way as for the first step, the first 42 days of observations are disregarded in the optimization of the tuning parameters. The benchmark models are initialized in the same manner, only using the spot price series instead of the residuals from the former step.

Again, least squares estimates of the parameters are sought yielding the ones summarized in the second column of Table II. The corresponding in- and out of-sample residual RMSEs and MAEs (in DKK and scaled) are given in the remaining columns along with that of the benchmark models.

Apart from the obvious fact that the proposed models drastically outperform the more naive benchmarks, the table reveals that the decaying unscaled performance between the training- and test periods may, to a certain degree, be explained by the somewhat greater volatility of the prices during the test period. This applies to both the full model and the intermediate step and further supports what already has been said about recursive robustification.

The forecasting skill of the Holt-Winters and the RLS-AR models applied to the residuals of the non-parametric model is also clearly superior to that of the models applied to the spot prices directly. As shown in Figure 4, which plots the hourly residual RMSE for the bottom four models in Table II, the superiority is a result of the two step models consistently outperforming the other models almost throughout the entire day. It is only in the first hour of the day that the benchmark RLS-AR model performs similarly to the two-step ones. This is because the forecasts for these hours are the ones with the shortest lead times and thus based on very recent observations.

The performance varies somewhat between hours but seems to coincide with the severity of the price spikes that occurred during the test period. This can be seen on the box plot in Figure 5 where the spot price distribution within each hour of the day is illustrated. The hours with the highest RMSE are among the ones when the most extreme prices occur.

The model's forecasting skill during normal weekdays, weekends and public holidays is listed in Table III. In terms of the unscaled measures, the performance seems to vary quite substantially between different types of days. However, the performance measures relative to that of the persistence forecast reveal that much of this variation is due to the alternating price volatility during the different day types. One has to bear in mind though that the unbalanced sample sizes between the different categories make this kind of comparison unreliable. That is, the small number of holidays makes performance

TABLE II
RMSE FOR THE DAY-AHEAD FORECASTS

Model		Parameters	In-sample $RMS(S)E$ [DKK]	In-sample $RMS(S)E$ $[-]$	In-sample $MA(S)E$ [DKK]	In-sample $MA(S)E$ $[-]$	Out-of-sample $RMS(S)E$ [DKK]	Out-of-sample $RMS(S)E$ $[-]$	Out-of-sample $MA(S)E$ [DKK]	Out-of-sample $MA(S)E$ $[-]$
Bench-marks	Period Mean	—	75.61	0.981	45.50	1.132	95.45	1.268	69.75	1.456
	Daily persistence	—	77.02	1.000	40.22	1.000	75.30	1.000	47.90	1.000
	Weekly persistence	—	74.37	0.965	41.69	1.037	79.83	1.060	52.42	1.094
	ARIMAX	—	69.20	0.898	35.93	0.894	64.21	0.853	43.26	0.903
	RLS-AR	$[\lambda, \tau] = [0.9889, 92.78]$	50.03	0.725	31.27	0.808	55.87	0.742	37.73	0.788
	Holt-Winters	$[\alpha_\mu, \alpha_D, \alpha_W, \tau] = [0.0116, 0.0903, 0.1009, 112.39]$	52.88	0.756	31.73	0.820	57.81	0.768	40.34	0.842
Two step	RLS-AR	$[\lambda, \tau] = [0.9915, 240.63]$	47.55	0.680	27.95	0.722	48.15	0.640	32.77	0.684
	Holt-Winters	$[\alpha_\mu, \alpha_D, \tau] = [0.0042, 0.1245, 32.98]$	46.81	0.669	27.23	0.704	49.07	0.652	33.66	0.703

TABLE III
OUT-OF-SAMPLE PERFORMANCE OF THE PROPOSED MODEL DURING WEEKDAYS, WEEKENDS AND HOLIDAYS SEPARATELY

Model	Day Type	RMS(S)E [DKK]	RMS(S)E [-]	MA(S)E [DKK]	MA(S)E [-]
RLS-AR	Weekdays	44.94	0.637	31.03	0.696
	Weekends	51.74	0.635	34.90	0.654
	Holidays	73.79	0.683	50.52	0.715
Holt-Winters	Weekdays	44.39	0.6287	30.67	0.6874
	Weekends	52.95	0.650	35.86	0.672
	Holidays	74.54	0.690	52.05	0.737

assessment during these days vulnerable for any extraordinary circumstances.

As previously mentioned, the explanatory power of the input forecasts used in the model's first step is owed to their reflection of the volumes bid to the market. The forecasting skill of the model can therefore be expected to be affected by how closely the input forecasts are related to the volume cleared on the market. Increased quality of the input forecast will therefore not necessarily improve the price forecasts unless resemblance with the bidding behavior is increased as well.

Overall, there seems to be little skill difference between the residual Holt-Winters and the residual RLS-AR models. In light of the fact that both mainly rely on the same information from the past this is understandable. Since both models are relatively easy to implement, choosing one out of the two thus comes down to personal preferences of the one implementing the model. The Holt-Winters model has the advantage though that price spikes are less likely to be reflected in forecasts for the following days since it does not explicitly use previous values for prediction. For the same reason, the Holt-Winters model is more robust operationally since missing observations will only affect the model update but will not prevent predictions from being issued.

V. CONCLUSIONS AND FUTURE WORK

A two step methodology for day-ahead forecasting of electricity spot prices has been presented. Whereas the first step accounts for the prices' dependence on forecast load and wind power production, the second step accommodates autocorrelation and seasonalities. The time-adaptive version of the model was shown to comfortably outperform its time-invariant counterpart. Hence, adaptive parameter estimation must be concluded to be relevant for the modeling of this phenomena. In terms of forecasting skill the interest of employing the robust approach is less obvious. However, the robust estimation protects the model's parameters from abrupt changes, caused by few excessive spikes. Thereby the model is enabled to follow the progress of the average prices more closely without manual inference. The time-varying price volatility suggest though that robustification should be made recursive.

Out-of-sample empirical results, obtained by mimicking practical circumstances, indicate that the model is well suited for practical use - both in terms of methodology and forecasting skill. In order to obtain complete forecasts of the electricity spot prices, the model here presented should be accompanied by a model for prediction intervals. Given the results of [3], such intervals would most likely be conditional upon fundamental factors, e.g. forecast wind power production and load. Whether modeling of higher order moments also requires time-adaptivity will be an interesting question to answer. Given the characteristics of the prices however, time-adaptivity is likely to be as essential in such models as it is here.

Even though the share of wind power in the generation portfolio is relatively large in DK-1, accounting for predicted wind power production is likely to be beneficial in other markets as well. For instance the findings of [5] hint that the methodology presented here could be successfully applied to the Spanish case. Here the fundamental difference between wind power and conventional power plants plays an essential role. In addition, the EU's target of having 20% of its energy consumption produced by renewable sources by 2020 and similar initiatives in the USA imply that price forecasting methods accounting for wind power production and renewable energy sources in general may have a more widespread applicability in the near future. In this context the inclusion of e.g. solar and wave power in the model parallel to their emergence would be interesting. Although both theoretically possible and not hard to implement, the inclusion of 1 or 2 more variables in the model calls for a more cautiously chosen variables or merger of them in order to ensure frequent enough updates of the parameters in all fitting points.

Acknowledgment

The authors would like to express their gratitute to the editors and the anonymous reviewers for providing useful comments that improved the paper substantially. Tryggvi Jónsson is partially sponsored by the Danish Agency for Science, Technology and Innovation which is hereby acknowledged. Pierre Pinson was partly supported by the European Commission through the EU FP6 project Anemos.plus (contract No 038692) and by the Danish Council for Strategic Research, Technology and Production through the Ensymora project (10-093904/DSF).

References

[1] I. Vehviläinen and T. Pyykkönen, "Stochastic factor model for electricity spot price - the case of the nordic market," *Energy Economics*, vol. 27(2), pp. 351–367, 2005.

[2] S.-E. Fleten and S. W. Wallace, "Power scheduling with forward contracts," in *Proceedings of the Nordic MPS*, Molde, Norway, 1998.

[3] T. Jónsson, P. Pinson, and H. Madsen, "On the market impact of wind energy forecasts," *Energy Economics*, vol. 32(2), pp. 313–320, 2010.

[4] P. Giabardo, M. Zugno, P. Pinson, and H. Madsen, "Feedback, competition and stochasticity in a day ahead electricity market," *Energy Economics*, vol. 32(2), pp. 292–301, 2010.

[5] J. M. Morales, A. J. Conejo, and J. Pérez-Ruiz, "Simulating the impact of wind production on locational marginal prices," *IEEE Transactions on Power Systems*, vol. 26, pp. 820–828, 2011.

[6] K. Skytte, "The regulating power market on the nordic power exchange nord pool: an econometric analysis," *Energy Economics*, vol. 21(4), pp. 295–308, 1999.

[7] P. E. Morthorst, "Wind power and the conditions at a liberalized power market," *Wind Energy*, vol. 6(3), pp. 297–308, 2003.

[8] A. J. Conejo, J. Contreras, R. Espínola, and M. A. Plazas, "Forecasting electricity prices for a day-ahead pool-based electric energy market," *International Journal of Forecasting*, vol. 21(3), pp. 435–462, 2005.

[9] A. Panagiotelis and M. Smith, "Bayesian density forecasting of intraday electricity prices using multivariate skew t distributions," *International Journal of Forecasting*, vol. 24(4), pp. 710–727, 2008.

[10] P. Kosater and K. Mosler, "Can Markov regime-switching models improve power-price forecasts? Evidence from German daily power prices," *Applied Energy*, vol. 83(9), pp. 943–958, 2006.

[11] M. Sewalt and C. de Jong, "Negative prices in electricity markets," *Commodities Now. Available at http://www.erasmusenergy.com/articles/91/1/Negative-prices-in-electricity-markets/Page1.html*, 2003.

[12] A. Boogert and D. Dupont, "On the effectiveness of the anti-gaming policy between the day-ahead and real-time electricity markets in the netherlands," *Energy Economics*, vol. 27(5), pp. 752–770, 2005.

[13] J. W. Taylor and P. E. McSharry, "Short-term load forecasting methods: An evaluation based on european data," *IEEE Transactions on Power Systems*, vol. 22(4), pp. 2213–2219, 2007.

[14] N. V. Karakatsani and D. W. Bunn, "Forecasting electricity prices: The impact of fundamentals and time-varying coefficients," *International Journal of Forecasting*, vol. 24(4), pp. 764–785, 2008.

[15] L. Wu and M. Shahidehpour, "A hybrid model for day-ahead price forecasting," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1519–1530, 2010.

[16] X. Lu, Z. Y. Dong, and X. Li, "Electricity market price spike forecast with data mining techniques," *Electric Power Systems Research*, vol. 73, no. 1, pp. 19–29, 2005.

[17] N. Amjady and F. Keynia, "Electricity market price spike analysis by a hybrid data model and feature selection technique," *Electric Power Systems Research*, vol. 80, no. 3, pp. 318–327, 2010.

[18] H. Higgs, "Modelling price and volatility inter-relationships in the Australian wholesale spot electricity markets," *Energy Economics*, vol. 31, no. 5, pp. 748–756, 2009.

[19] Nord Pool Spot AS. (2012, Jan.) Website. [Online]. Available: http://www.nordpoolspot.com

[20] Energinet.dk. (2012, Jan.) Website. [Online]. Available: http://www.energinet.dk

[21] T. S. Nielsen, H. A. Nielsen, and H. Madsen, "Prediction of wind power using time-varying coefficient functions," in *15th IFAC World Congress*, Barcelona, Spain, 2002.

[22] Nord Pool Spot AS. (2010, Nov.) Production split. [Online]. Available: http://www.nordpoolspot.com/reports/Production_split/

[23] B. Jonsson, "Prediction with a linear regression model and errors in a regressor," *International Journal of Forecasting*, vol. 10(4), pp. 549–555, 1994.

[24] W. S. Cleveland and S. J. Devlin, "Locally weighted regression: An approach to regression analysis by local fitting," *Journal of the American Statistical Association*, vol. 83(403), pp. 596–610, 1988.

[25] H. A. Nielsen, T. S. Nielsen, A. K. Joensen, H. Madsen, and J. Holst, "Tracking time-varying coefficient functions," *International Journal of Adaptive Control and Signal Processing*, vol. 14(8), pp. 813–828, 2000.

[26] P. Pinson, H. A. Nielsen, and H. Madsen, "Robust estimation of time-varying coefficient functions - application to the modeling of wind power production," Department of Informatics and Mathematical Modelling, Technical University of Denmark, Tech. Rep., 2007.

[27] P. J. Huber, *Robust Statistics*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 1981.

[28] H. Madsen, *Time Series Analysis*. London, UK: Chapman & Hall / CRC, 2008.

[29] G. Wahba, *Spline Models for Observational Data*. Philadelphia, PA, USA: SIAM, 1990.

[30] P. R. Winters, "Forecasting sales by exponentially weighted moving averages," *Management Science*, vol. 6(3), pp. 324–342, 1960.

[31] J. W. Taylor, "Short-term electricity demand forecasting using double seasonal exponential smoothing," *Journal of Operational Research Society*, vol. 54, pp. 799–805, 2003.

[32] R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose, "A state space framework for automatic forecasting using exponential smoothing methods," *International Journal of Forecasting*, vol. 18(3), pp. 439–454, 2002.

[33] S. Gelper, R. Fried, and C. Croux, "Robust forecasting with exponential and holt-winters smoothing," *Journal of Forecasting*, vol. 29(3), pp. 285–300, 2010.

[34] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, 2006.

[35] F. J. Nogales, J. Contreras, A. J. Conejo, and R. Espínola, "Forecasting next-day electricity prices by time series models," *IEEE Transactions on Power Systems*, vol. 17, no. 2, pp. 342–348, 2002.

[36] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning - Data Mining, Inference and Prediction*. New York, USA: Springer, 2001.

[37] K. Sejling, H. Madsen, J. Holst, U. Holst, and J.-E. Englund, "Methods for recursive robust estimation of ar parameters," *Computational Statistics & Data Analysis*, vol. 17, pp. 509–536, 1994.